Optimal Weight Choice for Frequentist Model Average Estimators

Hua LIANG, Guohua ZOU, Alan T. K. WAN, and Xinyu ZHANG

Supplementary Material

B.1 Derivation of $\Psi(\hat\theta, \hat\sigma^2)$ in Theorem 1

Along the lines of Carter, Srivastava, Srivastava and Ullah (1990) and Giles and Srivastava (1991), it can be shown that if $\Phi(\hat\theta, \hat\sigma^2)$ is a matrix differentiable with respect to $\hat\sigma^2$, then

$$E_{\hat\sigma^2}\left\{\hat\sigma^2\Phi(\hat\theta,\hat\sigma^2)\right\} = \sigma^2 E_{\hat\sigma^2}\left\{\Phi(\hat\theta,\hat\sigma^2)\right\} + \frac{2\sigma^2}{n-k-m}E_{\hat\sigma^2}\left\{\hat\sigma^2\frac{\partial\Phi(\hat\theta,\hat\sigma^2)}{\partial\hat\sigma^2}\right\}, \qquad \text{(S.1)}$$

provided that the expectations on both sides of (S.1) exist.

In fact, it is readily seen that

$$\begin{aligned}\int_0^\infty \Phi(\hat\theta,t)t^{(n-k-m)/2}e^{-(n-k-m)t/(2\sigma^2)}dt &= \frac{1}{C_0}\int_0^\infty t\Phi(\hat\theta,t)\cdot C_0 t^{(n-k-m)/2-1}e^{-(n-k-m)t/(2\sigma^2)}dt \\ &= \frac{1}{C_0}E_{\hat\sigma^2}\left\{\hat\sigma^2\Phi(\hat\theta,\hat\sigma^2)\right\}, \qquad \text{(S.2)}\end{aligned}$$

and

$$\begin{aligned}&\int_0^\infty \frac{\partial}{\partial t}\left\{\Phi(\hat\theta,t)t^{(n-k-m)/2}e^{-(n-k-m)t/(2\sigma^2)}\right\}dt \\ =& \int_0^\infty\left\{\frac{\partial\Phi(\hat\theta,t)}{\partial t}t^{(n-k-m)/2}e^{-(n-k-m)t/(2\sigma^2)} + \frac{n-k-m}{2}\Phi(\hat\theta,t)t^{(n-k-m)/2-1}e^{-(n-k-m)t/(2\sigma^2)}\right. \\ &\left. -\frac{n-k-m}{2\sigma^2}\Phi(\hat\theta,t)t^{(n-k-m)/2}e^{-(n-k-m)t/(2\sigma^2)}\right\}dt \\ =& \frac{1}{C_0}E_{\hat\sigma^2}\left\{\hat\sigma^2\frac{\partial\Phi(\hat\theta,\hat\sigma^2)}{\partial\hat\sigma^2} + \frac{n-k-m}{2}\Phi(\hat\theta,\hat\sigma^2) - \frac{n-k-m}{2\sigma^2}\hat\sigma^2\Phi(\hat\theta,\hat\sigma^2)\right\}. \qquad \text{(S.3)}\end{aligned}$$

Note that the elements of (S.2) and (S.3) are finite. Thus, we have

$$\lim_{t\to\infty}\left\{\Phi(\hat\theta,t)t^{(n-k-m)/2}e^{-(n-k-m)t/(2\sigma^2)}\right\} = 0.$$

Similarly,

$$\lim_{t\to 0}\left\{\Phi(\hat\theta,t)t^{(n-k-m)/2}e^{-(n-k-m)t/(2\sigma^2)}\right\} = 0.$$

Combining these two formulae, equation (S.1) is obvious from (S.3).

Now, let

$$\Phi(\hat{\theta}, \hat{\sigma}^2) + \frac{2}{n-k-m} \cdot \hat{\sigma}^2 \frac{\partial \Phi(\hat{\theta}, \hat{\sigma}^2)}{\partial \hat{\sigma}^2} = \Psi_1(\hat{\theta}, \hat{\sigma}^2).$$

Then a solution to the above differential equation is given by

$$\Phi_0(\hat{\theta}, \hat{\sigma}^2) = \frac{n-k-m}{2} (\hat{\sigma}^2)^{-(n-k-m)/2} \int_0^{\hat{\sigma}^2} t^{(n-k-m)/2-1} \Psi_1(\hat{\theta}, t) dt.$$

Applying these results in equation (S.1), we may write $\sigma^2 E_{\hat{\sigma}^2} \left\{ \Psi_1(\hat{\theta}, \hat{\sigma}^2) \right\} = E_{\hat{\sigma}^2} \left\{ \hat{\sigma}^2 \Phi_0(\hat{\theta}, \hat{\sigma}^2) \right\}$. Thus, we take $\Psi(\hat{\theta}, \hat{\sigma}^2) = \hat{\sigma}^2 \Phi_0(\hat{\theta}, \hat{\sigma}^2)$, which is precisely equation (5).

B.2 Numerical comparison of $\hat{\sigma}^2 \Psi_1(\hat{\theta}, \hat{\sigma}^2)$ and $\Psi(\hat{\theta}, \hat{\sigma}^2)$

The aim of this numerical exercise is to examine how well the values of $\hat{\sigma}^2 \Psi_1(\hat{\theta}, \hat{\sigma}^2)$ accord with those of $\Psi(\hat{\theta}, \hat{\sigma}^2)$. Our set-up is the same as that used for Example 1 in the simulation study of Section 4. We consider the S-AIC, S-BIC, S-RMS and smoothed generalized cross-validation (S-GCV) weights, and let n = 100, 900, α = 0.1, 0.9, and R^2 vary between 0.1 and 0.9.

Let Γ be an $m \times m$ matrix with the ij^{th} element being

$$\Gamma_{ij} = \left| \frac{\hat{\sigma}^2 \Psi_1(\hat{\theta}, \hat{\sigma}^2)_{ij} - \Psi(\hat{\theta}, \hat{\sigma}^2)_{ij}}{\Psi(\hat{\theta}, \hat{\sigma}^2)_{ij}} \right|,$$

where $\Psi_1(\hat{\theta}, \hat{\sigma}^2)_{ij}$ and $\Psi(\hat{\theta}, \hat{\sigma}^2)_{ij}$ are the ij^{th} element of $\Psi_1(\hat{\theta}, \hat{\sigma}^2)$ and $\Psi(\hat{\theta}, \hat{\sigma}^2)$ respectively, $i = 1, \cdots, m$, $j = 1, \cdots, m$. So, Γ_{ij} measures the relative discrepancy when $\hat{\sigma}^2 \Psi_1(\hat{\theta}, \hat{\sigma}^2)_{ij}$ is used as a proxy for $\Psi(\hat{\theta}, \hat{\sigma}^2)_{ij}$. A zero or near-zero value of Γ_{ij} shows that $\hat{\sigma}^2 \Psi_1(\hat{\theta}, \hat{\sigma}^2)_{ij}$ is a good proxy for $\Psi(\hat{\theta}, \hat{\sigma}^2)_{ij}$, and vice versa. Now, to facilitate comparisons, let us define the average relative discrepancy as $\gamma = \frac{1}{m^2} \sum\limits_{i=1}^{m} \sum\limits_{j=1}^{m} \Gamma_{i,j}$. Figure B.1 gives the values of $\Delta = \sum\limits_{r=1}^{M} \gamma^{(r)}/M$ for the different cases, with $M = 100$ being the total number of replications and r denoting the r^{th} replication.

Figure B.1 shows that $\hat{\sigma}^2 \Psi_1(\hat{\theta}, \hat{\sigma}^2)$ is generally a good proxy for $\Psi(\hat{\theta}, \hat{\sigma}^2)$, with the average relative discrepancy being near 0 for all cases considered.

B.3 Proof of (A.24)

Using (A.20), (A.22), and (A.23), it can be shown that for any $\delta > 0$,

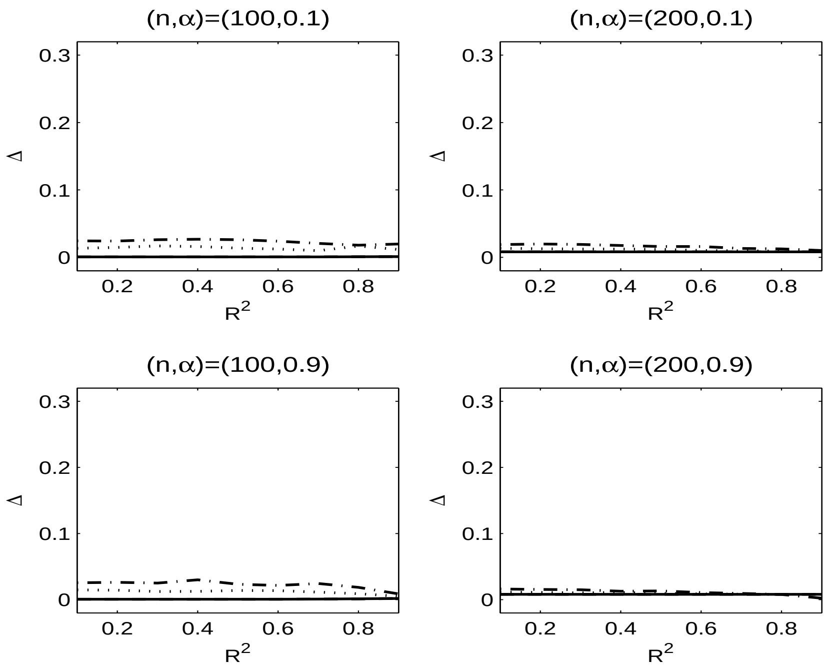

Figure B.1: *Comparison of $\hat{\sigma}^2\Psi_1(\hat{\theta},\hat{\sigma}^2)$ and $\Psi(\hat{\theta},\hat{\sigma}^2)$. The dash-dotted, dotted, dashed, and solid lines correspond to S-BIC, S-AIC, S-RMS and S-GCV weights respectively.*

$$
\begin{aligned}
&\Pr\left\{\left|\frac{\inf_{(a,b,c)\in\mathcal{D}_0} L_n(\lambda(a,b,c))}{L_n(\lambda(\hat{a},\hat{b},\hat{c}))}-1\right|>\delta\right\}=\Pr\left\{\frac{L_n(\lambda(\hat{a},\hat{b},\hat{c}))-\inf_{(a,b,c)\in\mathcal{D}_0} L_n(\lambda(a,b,c))}{L_n(\lambda(\hat{a},\hat{b},\hat{c}))}>\delta\right\}\\
&=\Pr\left\{\frac{\inf_{(a,b,c)\in\mathcal{D}_0}\left(L_n(\lambda(a,b,c))+t_n(a,b,c)\right)-t_n(\hat{a},\hat{b},\hat{c})-\inf_{(a,b,c)\in\mathcal{D}_0} L_n(\lambda(a,b,c))}{L_n(\lambda(\hat{a},\hat{b},\hat{c}))}>\delta\right\}\\
&\leq\Pr\left\{\frac{L_n(\lambda(a_n,b_n,c_n))+t_n(a_n,b_n,c_n)-t_n(\hat{a},\hat{b},\hat{c})-L_n(\lambda(a_n,b_n,c_n))+\vartheta_n}{L_n(\lambda(\hat{a},\hat{b},\hat{c}))}>\delta\right\}\\
&\leq\Pr\left\{\frac{|t_n(a_n,b_n,c_n)|}{L_n(\lambda(\hat{a},\hat{b},\hat{c}))}+\frac{|t_n(\hat{a},\hat{b},\hat{c})|}{L_n(\lambda(\hat{a},\hat{b},\hat{c}))}+\frac{\vartheta_n}{L_n(\lambda(\hat{a},\hat{b},\hat{c}))}>\delta\right\}\\
&\leq 2\Pr\left\{\frac{\sup_{(a,b,c)\in\mathcal{D}_0}|t_n(a,b,c)|}{\inf_{(a,b,c)\in\mathcal{D}_0} L_n(\lambda(a,b,c))}>\frac{\delta}{3}\right\}+\Pr\left\{\frac{\vartheta_n}{\inf_{(a,b,c)\in\mathcal{D}_0} L_n(\lambda(a,b,c))}>\frac{\delta}{3}\right\}\\
&\leq 2\Pr\left\{\frac{\sup_{(a,b,c)\in\mathcal{D}_0}|t_n(a,b,c)|}{\inf_{w\in\mathcal{W}} L_n(w)}>\frac{\delta}{3}\right\}+\Pr\left\{\frac{\vartheta_n}{\inf_{w\in\mathcal{W}} L_n(w)}>\frac{\delta}{3}\right\}\\
&=2\Pr\left\{\frac{\sup_{(a,b,c)\in\mathcal{D}_0}|t_n(a,b,c)|}{\xi_n}\cdot\frac{\xi_n}{\inf_{w\in\mathcal{W}}\left(R_n(w)+u_n(w)\right)}>\frac{\delta}{3}\right\}
\end{aligned}
$$

$$
\begin{aligned}
&+\Pr\left\{\frac{\vartheta_n}{\xi_n}\cdot\frac{\xi_n}{\inf_{w\in\mathcal{W}}(R_n(w)+u_n(w))}>\frac{\delta}{3}\right\}\\
&=2\Pr\left\{\left(\frac{\sup_{(a,b,c)\in\mathcal{D}_0}|t_n(a,b,c)|}{\xi_n}\cdot\frac{\xi_n}{\inf_{w\in\mathcal{W}}(R_n(w)+u_n(w))}>\frac{\delta}{3}\right)\bigcap\left(\left|\inf_{w\in\mathcal{W}}u_n(w)\right|<\inf_{w\in\mathcal{W}}R_n(w)\right)\right\}\\
&+2\Pr\left\{\left(\frac{\sup_{(a,b,c)\in\mathcal{D}_0}|t_n(a,b,c)|}{\xi_n}\cdot\frac{\xi_n}{\inf_{w\in\mathcal{W}}(R_n(w)+u_n(w))}>\frac{\delta}{3}\right)\bigcap\left(\left|\inf_{w\in\mathcal{W}}u_n(w)\right|\geq\inf_{w\in\mathcal{W}}R_n(w)\right)\right\}\\
&+\Pr\left\{\left(\frac{\vartheta_n}{\xi_n}\cdot\frac{\xi_n}{\inf_{w\in\mathcal{W}}(R_n(w)+u_n(w))}>\frac{\delta}{3}\right)\bigcap\left(\left|\inf_{w\in\mathcal{W}}u_n(w)\right|<\inf_{w\in\mathcal{W}}R_n(w)\right)\right\}\\
&+\Pr\left\{\left(\frac{\vartheta_n}{\xi_n}\cdot\frac{\xi_n}{\inf_{w\in\mathcal{W}}(R_n(w)+u_n(w))}>\frac{\delta}{3}\right)\bigcap\left(\left|\inf_{w\in\mathcal{W}}u_n(w)\right|\geq\inf_{w\in\mathcal{W}}R_n(w)\right)\right\}\\
&\leq 2\Pr\left\{\frac{\sup_{(a,b,c)\in\mathcal{D}_0}|t_n(a,b,c)|}{\xi_n}\cdot\frac{\xi_n}{\inf_{w\in\mathcal{W}}R_n(w)+\inf_{w\in\mathcal{W}}u_n(w)}>\frac{\delta}{3}\right\}\\
&+\Pr\left\{\frac{\vartheta_n}{\xi_n}\cdot\frac{\xi_n}{\inf_{w\in\mathcal{W}}R_n(w)+\inf_{w\in\mathcal{W}}u_n(w)}>\frac{\delta}{3}\right\}+3\Pr\left\{\left|\inf_{w\in\mathcal{W}}u_n(w)\right|\geq\inf_{w\in\mathcal{W}}R_n(w)\right\}\\
&=2\Pr\left\{\frac{\sup_{(a,b,c)\in\mathcal{D}_0}|t_n(a,b,c)|}{\xi_n}\cdot\frac{1}{1+\inf_{w\in\mathcal{W}}u_n(w)/\xi_n}>\frac{\delta}{3}\right\}\\
&+\Pr\left\{\frac{\vartheta_n}{\xi_n}\cdot\frac{1}{1+\inf_{w\in\mathcal{W}}u_n(w)/\xi_n}>\frac{\delta}{3}\right\}+3\Pr\left\{\frac{\left|\inf_{w\in\mathcal{W}}u_n(w)\right|}{\xi_n}\geq 1\right\}.
\end{aligned}
$$

Hence, (A.24) is established.

B.4 Proof of (A.28)

First, the risk of $\hat{\mu}_{(i)}$ is

$$E\|\hat{\mu}_{(i)}-\mu\|^2=\|A_i\mu\|^2+\sigma^2 q_i. \tag{S.4}$$

By Chebyshev's inequality, Theorem 2 of Whittle (1960) and (S.4), we observe, for any $\delta>0$, that

$$
\begin{aligned}
\Pr\left\{\sup_{w\in\mathcal{W}}|\mu' A(w)\varepsilon|/\xi_n>\delta\right\} &\leq \Pr\left\{\sup_{w\in\mathcal{W}}\sum_{i=1}^N w_i\,|\mu' A_i\varepsilon|>\delta\xi_n\right\}\\
&\leq \Pr\left\{\max_{1\leq i\leq N}|\mu' A_i\varepsilon|>\delta\xi_n\right\}
\end{aligned}
$$

$$
\begin{aligned}
&\leq \sum_{i=1}^{N} \Pr\left\{|\mu' A_i \varepsilon| > \delta \xi_n\right\} \\
&\leq \sum_{i=1}^{N} E\left\{\left(\mu' A_i \varepsilon\right)^2 \delta^{-2} \xi_n^{-2}\right\} \\
&\leq C_1 \delta^{-2} \xi_n^{-2} \sum_{i=1}^{N} \|A_i \mu\|^2 \\
&\leq C_2 \delta^{-2} \xi_n^{-2} \zeta_n, \qquad \text{(S.5)}
\end{aligned}
$$

where C_1 and C_2 are positive constants. Thus, from condition (22), we obtain (A.28).

B.5 Proof of (A.29)

By Chebyshev's inequality, Theorem 2 of Whittle (1960) and (S.4), we have

$$
\begin{aligned}
\Pr\left\{\sup_{w \in \mathcal{W}} \left|\varepsilon' T(w) \varepsilon - \sigma^2 w' q\right| / \xi_n > \delta\right\} &\leq \sum_{i=1}^{N} \Pr\left\{\left|\varepsilon' T_i \varepsilon - \sigma^2 q_i\right| > \delta \xi_n\right\} \\
&\leq \sum_{i=1}^{N} E\left\{\left(\varepsilon' T_i \varepsilon - \sigma^2 q_i\right)^2 \delta^{-2} \xi_n^{-2}\right\} \\
&\leq C_3 \delta^{-2} \xi_n^{-2} \sum_{i=1}^{N} q_i,
\end{aligned}
$$

where C_3 is a positive constant. Together with condition (22) and the fact that $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$, this leads to (A.29).

B.6 Proofs of (A.30) and (A.31)

First, note that for any $i, j \in \{1, ..., N\}$, $\|T_i A_j \mu\|^2 \leq \|A_j \mu\|^2$. Similar to the proof of (S.5), we have

$$
\begin{aligned}
\Pr\left\{\sup_{w \in \mathcal{W}} |\mu' A(w) T(w) \varepsilon| / \xi_n > \delta\right\} &\leq \Pr\left\{\sup_{w \in \mathcal{W}} \sum_{j=1}^{N} \sum_{i=1}^{N} w_j w_i |\mu' A_i T_j \varepsilon| > \delta \xi_n\right\} \\
&\leq \Pr\left\{\max_{1 \leq j \leq N} \max_{1 \leq i \leq N} |\mu' A_i T_j \varepsilon| > \delta \xi_n\right\} \\
&\leq \sum_{j=1}^{N} \sum_{i=1}^{N} \Pr\left\{|\mu' A_j T_i \varepsilon| > \delta \xi_n\right\} \\
&\leq \sum_{j=1}^{N} \sum_{i=1}^{N} E\left[(\mu' A_j T_i \varepsilon)^2 \delta^{-2} \xi_n^{-2}\right] \\
&\leq C_4 \delta^{-2} \xi_n^{-2} \sum_{j=1}^{N} \sum_{i=1}^{N} \|T_i A_j \mu\|^2 \\
&\leq C_4 \delta^{-2} \xi_n^{-2} \sum_{j=1}^{N} \sum_{i=1}^{N} \|A_j \mu\|^2 \\
&\leq C_5 \delta^{-2} \xi_n^{-2} \zeta_n,
\end{aligned}
$$

where C_4 and C_5 are positive constants. Hence, equation (A.30) is obtained from condition (22).

Likewise, recognizing that for any $i, j \in \{1, ..., N\}$, tr$(T_iT_j) \leq$ tr(T_j), we have

$$
\begin{aligned}
&\Pr\left\{\sup_{w\in\mathcal{W}}\left|\|T(w)\varepsilon\|^2 - \sigma^2\text{tr}\left\{T^2(w)\right\}\right|/\xi_n > \delta\right\}\\
&\leq \Pr\left\{\sup_{w\in\mathcal{W}}\sum\nolimits_{j=1}^N\sum\nolimits_{i=1}^N w_jw_i\left|\varepsilon' T_jT_i\varepsilon - \sigma^2\text{tr}(T_jT_i)\right| > \delta\xi_n\right\}\\
&\leq \Pr\left\{\max_{1\leq j\leq N}\max_{1\leq i\leq N}\left|\varepsilon' T_jT_i\varepsilon - \sigma^2\text{tr}(T_jT_i)\right| > \delta\xi_n\right\}\\
&\leq \sum\nolimits_{j=1}^N\sum\nolimits_{i=1}^N \Pr\left\{\left|\varepsilon' T_jT_i\varepsilon - \sigma^2\text{tr}(T_jT_i)\right| > \delta\xi_n\right\}\\
&\leq \sum\nolimits_{j=1}^N\sum\nolimits_{i=1}^N E\left[\left\{\varepsilon' T_jT_i\varepsilon - \sigma^2\text{tr}(T_jT_i)\right\}^2\delta^{-2}\xi_n^{-2}\right]\\
&\leq C_6\delta^{-2}\xi_n^{-2}\sum\nolimits_{j=1}^N\sum\nolimits_{i=1}^N \text{tr}(T_j^2T_i^2)\\
&= C_6\delta^{-2}\xi_n^{-2}\sum\nolimits_{j=1}^N\sum\nolimits_{i=1}^N \text{tr}(T_jT_i)\\
&\leq C_6\delta^{-2}\xi_n^{-2}\sum\nolimits_{j=1}^N\sum\nolimits_{i=1}^N q_i,
\end{aligned}
$$

where C_6 is a positive constant. Thus, by condition (22), (A.31) is also true.

B.7 Proof of the result on the upper bound of $\sum\limits_{\tau\in\mathcal{U}}\lambda_\tau(a,b,c)$ in Section 3.3

Recall that we assume $-\bar{c} \leq c \leq 0$, κ_1 and κ_2 exist such that $0 < \kappa_1 \leq \kappa_2 < \infty$ and $\kappa_1 \leq \hat{\sigma}_i^2/\sigma^2 \leq \kappa_2$ with probability one for any $i \in \{1, ..., N\}$, and $\bar{a}_1$, $\bar{a}_2$ and $\bar{b}$ exist such that $0 < \bar{a}_1 \leq a \leq \bar{a}_2 < \infty, 0 \leq b \leq \bar{b} < \infty$. Under these assumptions, it is seen that for any $\tau \in \mathcal{U}$ and $i \notin \mathcal{U}$,

$$
\begin{aligned}
\frac{\lambda_i(a,b,c)}{\lambda_\tau(a,b,c)} &= \frac{a^{q_i}(n-q_i)^b(\hat{\sigma}_i^2)^c}{a^{q_\tau}(n-q_\tau)^b(\hat{\sigma}_\tau^2)^c} \geq \min\left\{\bar{a}_1^{q_i-q_\tau}, \bar{a}_2^{q_i-q_\tau}\right\}\left(\frac{n-q_i}{n}\right)^{\bar{b}}\left(\frac{\hat{\sigma}_\tau^2}{\hat{\sigma}_i^2}\right)^{-c}\\
&\geq \min\left\{\bar{a}_1^{q_i-q_\tau}, \bar{a}_2^{q_i-q_\tau}\right\}\left(1-\frac{q_i}{q_i+1}\right)^{\bar{b}}\left(\frac{\kappa_1}{\kappa_2}\right)^{\bar{c}} \equiv v_{i,\tau} > 0
\end{aligned}
$$

with probability one. Let $v = \min\limits_{\tau\in\mathcal{U}, i\notin\mathcal{U}} v_{i,\tau} > 0$, $i^* \notin \mathcal{U}$, and $\#\mathcal{U}$ be the number of the elements in $\mathcal{U}$. Then, with probability one,

$$
\begin{aligned}
\frac{\sum_{\tau\in\mathcal{U}}\lambda_\tau(a,b,c)}{1-\sum_{\tau\in\mathcal{U}}\lambda_\tau(a,b,c)} &= \frac{\sum_{\tau\in\mathcal{U}}\lambda_\tau(a,b,c)}{\sum_{i\notin\mathcal{U}}\lambda_i(a,b,c)} \leq \frac{\sum_{\tau\in\mathcal{U}}\lambda_\tau(a,b,c)}{\lambda_{i^*}(a,b,c)}\\
&= \sum\nolimits_{\tau\in\mathcal{U}}\frac{\lambda_\tau(a,b,c)}{\lambda_{i^*}(a,b,c)} \leq \frac{\#\mathcal{U}}{v} \equiv v^* < \infty,
\end{aligned}
$$

which in turn implies

$$
\sum\nolimits_{\tau\in\mathcal{U}}\lambda_\tau(a,b,c) \leq \frac{v^*}{1+v^*} \equiv 1-\rho
$$

with probability one, where $0 < \rho < 1$.

B.8 *Proofs of the results related to the simple example in Section 3.3*

Note that there are two sub-models - one is the restricted model with X being the only regressor, and the other is the unrestricted model which is also the only unbiased model in the set-up. Let $\hat{\mu}_{(1)}$ and $\hat{\mu}_{(2)}$ be the estimators of μ based on the restricted and unrestricted models respectively. Recognizing that the vectors X and Z are orthogonal and $Z'Z = n/2$, it can be seen from (S.4) that

$$E\|\hat{\mu}_{(1)} - \mu\|^2 = \gamma^2 Z'(I_n - XX'/n)Z + \sigma^2 = \gamma^2 Z'Z + \sigma^2 = 0.01n/2 + \sigma^2$$

and

$$E\|\hat{\mu}_{(2)} - \mu\|^2 = 2\sigma^2.$$

Thus, $\zeta_n = 0.005n + \sigma^2$ when $n \geq 200\sigma^2$. From the third equality in formula (A.21) of the Appendix, we obtain

$$\begin{aligned} R_n(w) &= \|T(w)\mu - \mu\|^2 + \sigma^2 \text{tr}\{T^2(w)\} \\ &= (w_1, w_2) \begin{pmatrix} 0.005n + \sigma^2 & \sigma^2 \\ \sigma^2 & 2\sigma^2 \end{pmatrix} (w_1, w_2)' \\ &= w_1^2(0.005n + \sigma^2) + 2w_2^2\sigma^2 + 2w_1 w_2 \sigma^2. \end{aligned}$$

So, $\xi_n = \inf_{w \in \mathcal{W}} R_n(w) \geq \rho^2(0.005n + \sigma^2) = \rho^2 \zeta_n$ when $n \geq 200\sigma^2$.

B.9 *Other examples in which condition (22) is satisfied*

Here, in addition to the example shown in the paper, we provide two other examples in which condition (22) is satisfied.

Example B.9.1: Consider a nested model set-up as in Hansen (2007, Econometrica), and assume that condition (21) holds. A sufficient condition for (22) to hold is

$$n^{1/2} \left(\|A_\tau \mu\|^2 \right)^{-1} = o(1), \qquad \tau \notin \mathcal{U}. \tag{S.6}$$

In fact, from the third equality of (A.21), we have

$$\begin{aligned} R_n(w) &= \|T(w)\mu - \mu\|^2 + \sigma^2 \text{tr}\{T^2(w)\} \\ &= \|\sum_{i=1}^N w_i A_i \mu\|^2 + \sigma^2 \text{tr}\left(\sum_{i=1}^N w_i T_i \right)^2 \\ &= w'Q_1 w + \sigma^2 w' Q_2 w, \end{aligned}$$

with the ij^{th} element of Q_1 being $Q_{1,ij} = \|A_{\max\{i,j\}}\mu\|^2$, and the ij^{th} element of Q_2 being $Q_{2,ij} = q_{\min\{i,j\}}$. So, for any $w \in \mathcal{W}$,

$$R_n(w) \geq \rho^2 \min_{\tau \notin \mathcal{U}} \|A_\tau \mu\|^2 + \sigma^2 \min_{i \in \{1,...,N\}} q_i,$$

and

$$\zeta_n = \max_{i \in \{1,...,N\}} E\|\hat{\mu}_{(i)} - \mu\|^2 = \max_{i \in \{1,...,N\}} \left\{\|A_i\mu\|^2 + \sigma^2 q_i\right\} \leq \mu'\mu + \sigma^2 \max_{i \in \{1,...,N\}} q_i.$$

From the above two formulae and condition (S.6), we have

$$\begin{aligned} \xi_n^{-2}\zeta_n &\leq \frac{\mu'\mu + \sigma^2 \max_{i \in \{1,...,N\}} q_i}{\left\{\rho^2 \min_{\tau \in \mathcal{U}} \|A_\tau\mu\|^2 + \sigma^2 \min_{i \in \{1,...,N\}} q_i\right\}^2} \\ &= \left(\frac{n^{1/2}}{\rho^2 \min_{\tau \in \mathcal{U}} \|A_\tau\mu\|^2 + \sigma^2 \min_{i \in \{1,...,N\}} q_i}\right)^2 \frac{\mu'\mu + \sigma^2 \max_{i \in \{1,...,N\}} q_i}{n} \\ &\to 0. \end{aligned}$$

A situation under which condition (S.6) holds is where H has orthogonal columns with norm $\|h_j\|^2 \sim n$. In this case, for any $\tau \notin \mathcal{U}$,

$$\|A_\tau\mu\|^2 = \mu'\left(I_n - H_\tau(H_\tau' H_\tau)^{-1} H_\tau'\right)\mu = \bar{\theta}'\bar{H}_\tau'\bar{H}_\tau\bar{\theta} \sim n,$$

where $\bar{H}_\tau$ consists of the columns of H not in H_τ, and $\bar{\theta}$ is the coefficient vector corresponding to $\bar{H}_\tau$.

Example B.9.2: This example concerns a non-nested set-up. The data are generated from

$$y = \mu + \varepsilon = \beta l_n + \gamma_1 t_n + \gamma_2 h_n + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2 I_n),$$

where β and γ_1 are fixed positive constants, $\gamma_2 = 0$, and $l_n = (1, ..., 1)'$, $t_n = (1, 2, ..., n)'/n$ and $h_n = (0, ..., 0, 1, ..., 1)'$ are all $n \times 1$ vectors (it is assumed that when n is even, the vector h_n contains equal numbers of 0's and 1's; on the other hand, for odd n, the number of 0's in h_n is one less than that of 1's). Let t_n and h_n be auxiliary regressors, and we consider $N = 2^2 = 4$

sub-models with the regressor matrices in the four sub-models being $H_1 = l_n$, $H_2 = [l_n : t_n]$, $H_3 = [l_n : h_n]$ and $H_4 = [l_n : t_n : h_n]$ respectively. Clearly, the first and third sub-models are biased, while the second and fourth sub-models are unbiased.

Using the third equality of (A.21), we have

$$R_n(w) = w'Q_1 w + \sigma^2 w'Q_2 w, \tag{S.7}$$

where the ij^{th} elements of Q_1 and Q_2 are $Q_{1,ij} = \mu' A_i A_j \mu$ and $Q_{2,ij} = \mathrm{tr}(T_i T_j)$ respectively. Recall that the second and fourth sub-models are unbiased, so we have $A_2\mu = 0$ and $A_4\mu = 0$, and thus $Q_{1,12} = Q_{1,14} = Q_{1,22} = Q_{1,23} = Q_{1,24} = Q_{1,34} = Q_{1,44} = 0$. In addition, the first sub-model is nested within the third, so $Q_{1,13} = \mu' A_1 A_3 \mu = \mu' A_3 \mu = Q_{1,33} \geq 0$. Furthermore,

$$\begin{aligned} Q_{1,11} &= \mu' A_1 \mu = \gamma_1^2 t_n'(I_n - T_1)t_n = \gamma_1^2 \left(t_n' t_n - t_n' l_n l_n' t_n / n\right) \\ &= \gamma_1^2 \left(\frac{n(n+1)(2n+1)}{6n^2} - \frac{n^2(n+1)^2}{4n^3}\right) \sim n, \end{aligned} \tag{S.8}$$

and

$$\begin{aligned}
& Q_{1,33} = \mu' A_3 \mu = \gamma_1^2 t_n'(I_n - T_3)t_n = \gamma_1^2 \left(t_n' t_n - t_n'[l_n : h_n] \left([l_n : h_n]'[l_n : h_n]\right)^{-1} [l_n : h_n]' t_n\right) \\
&= \begin{cases} \gamma_1^2 \left(\frac{(n+1)(2n+1)}{6n} - \left[\frac{n+1}{2}, \frac{3n+2}{8}\right] \begin{bmatrix} n & \frac{n}{2} \\ \frac{n}{2} & \frac{n}{2} \end{bmatrix}^{-1} \left[\frac{n+1}{2}, \frac{3n+2}{8}\right]'\right), \text{ for even } n \\ \gamma_1^2 \left(\frac{(n+1)(2n+1)}{6n} - \left[\frac{n+1}{2}, \frac{(n+1)(3n+1)}{8n}\right] \begin{bmatrix} n & \frac{n+1}{2} \\ \frac{n+1}{2} & \frac{n+1}{2} \end{bmatrix}^{-1} \left[\frac{n+1}{2}, \frac{(n+1)(3n+1)}{8n}\right]'\right), \text{ for odd } n \end{cases} \\
&= \begin{cases} \gamma_1^2 \left(\frac{(n+1)(2n+1)}{6n} - \frac{2}{n} \left[\frac{n+1}{2}, \frac{3n+2}{8}\right] \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}^{-1} \left[\frac{n+1}{2}, \frac{3n+2}{8}\right]'\right), \text{ for even } n \\ \gamma_1^2 \left(\frac{(n+1)(2n+1)}{6n} - \frac{2}{n+1} \left[\frac{n+1}{2}, \frac{(n+1)(3n+1)}{8n}\right] \begin{bmatrix} \frac{2n}{n+1} & 1 \\ 1 & 1 \end{bmatrix}^{-1} \left[\frac{n+1}{2}, \frac{(n+1)(3n+1)}{8n}\right]'\right), \text{ for odd } n \end{cases} \\
&= \begin{cases} \gamma_1^2 \left(\frac{(n+1)(2n+1)}{6n} - \frac{2}{n} \left[\frac{n+1}{2}, \frac{3n+2}{8}\right] \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \left[\frac{n+1}{2}, \frac{3n+2}{8}\right]'\right), \text{ for even } n \\ \gamma_1^2 \left(\frac{(n+1)(2n+1)}{6n} - \frac{2}{n+1} \left[\frac{n+1}{2}, \frac{(n+1)(3n+1)}{8n}\right] \begin{bmatrix} \frac{n+1}{n-1} & -\frac{n+1}{n-1} \\ -\frac{n+1}{n-1} & \frac{2n}{n-1} \end{bmatrix} \left[\frac{n+1}{2}, \frac{(n+1)(3n+1)}{8n}\right]'\right), \text{ for odd } n \end{cases} \\
&= \begin{cases} \gamma_1^2 \left(\frac{(n+1)(2n+1)}{6n} - \frac{(n+1)^2}{2n} - \frac{(3n+2)^2}{16n} + \frac{(n+1)(3n+2)}{4n}\right), \text{ for even } n \\ \gamma_1^2 \left(\frac{(n+1)(2n+1)}{6n} - \frac{(n+1)^2}{2(n-1)} - \frac{(n+1)(3n+1)^2}{16n(n-1)} + \frac{(n+1)^2(3n+1)}{4n(n-1)}\right), \text{ for odd } n \end{cases} \\
& \sim n.
\end{aligned} \tag{S.9}$$

Combining (S.7), (S.8), and (S.9), and noting that $Q_{2,ij}$ is bounded and $w_1 + w_3 \geq \rho$, we have $\zeta_n \sim n$ and $\xi_n \sim n$. Thus, condition (22) is satisfied.

B.10 Further results for simulation Example 1 in Section 4

Figure B.2 depicts results for parameter settings of simulation Example 1 not covered in Figure 4. As can be seen from Figure B.2, although there are exceptions, over a relatively large region of the parameter space, the OPT estimator is the most preferred estimator while the S-BIC estimator is generally the worst. These observations are consistent with those observed in Figure 4.

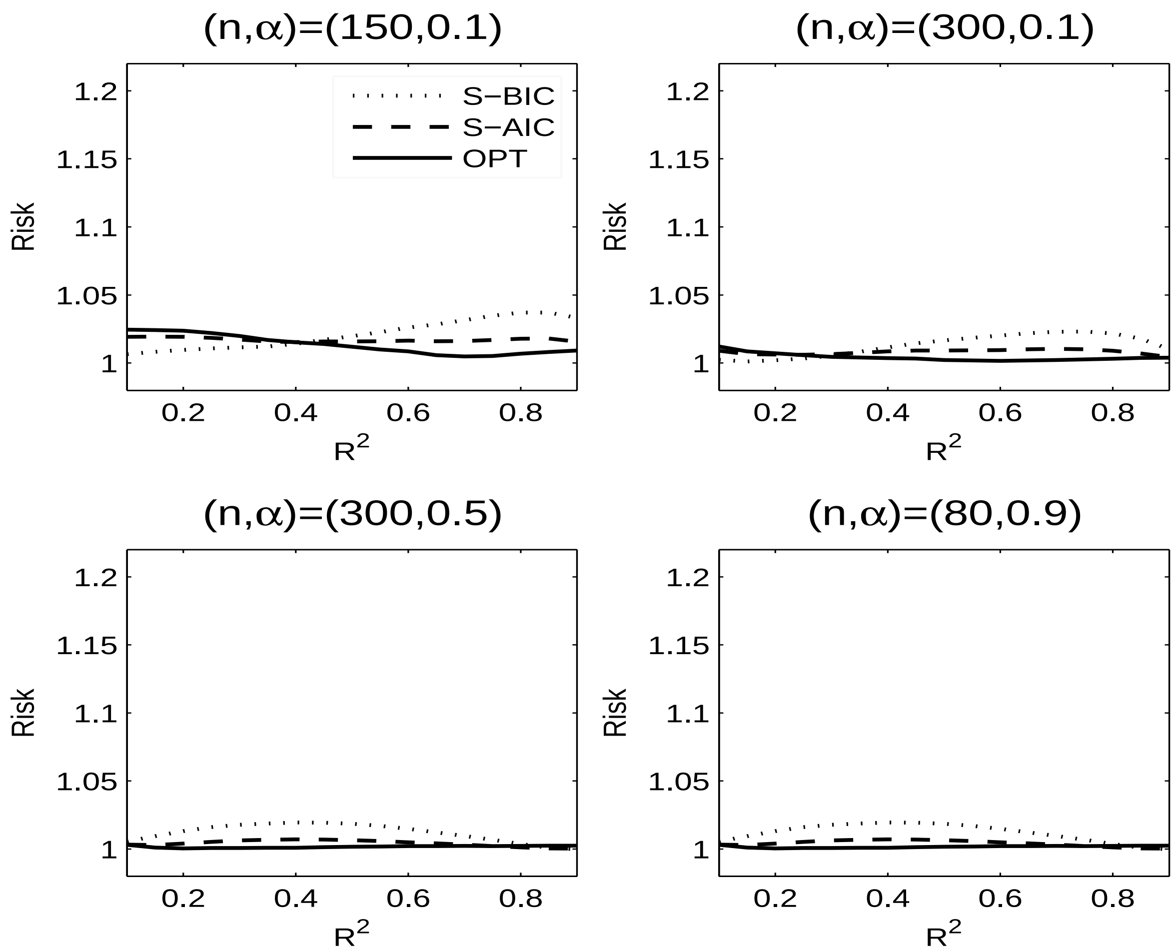

Figure B.2: Results for Example 1: risk comparisons under $L^{(2)}$ loss.

B.11 Further results for simulation Example 2 in Section 4

Figure B.3 provides results for other parameter settings of Example 2 not covered in Figure 5. As can be seen from Figure B.3, while there are exceptions, over a relatively large region of the parameter space, the OPT estimator is superior to the MMA estimator, and this superiority is especially noticeable when n is large. These observations are consistent with those observed in Figure 5.

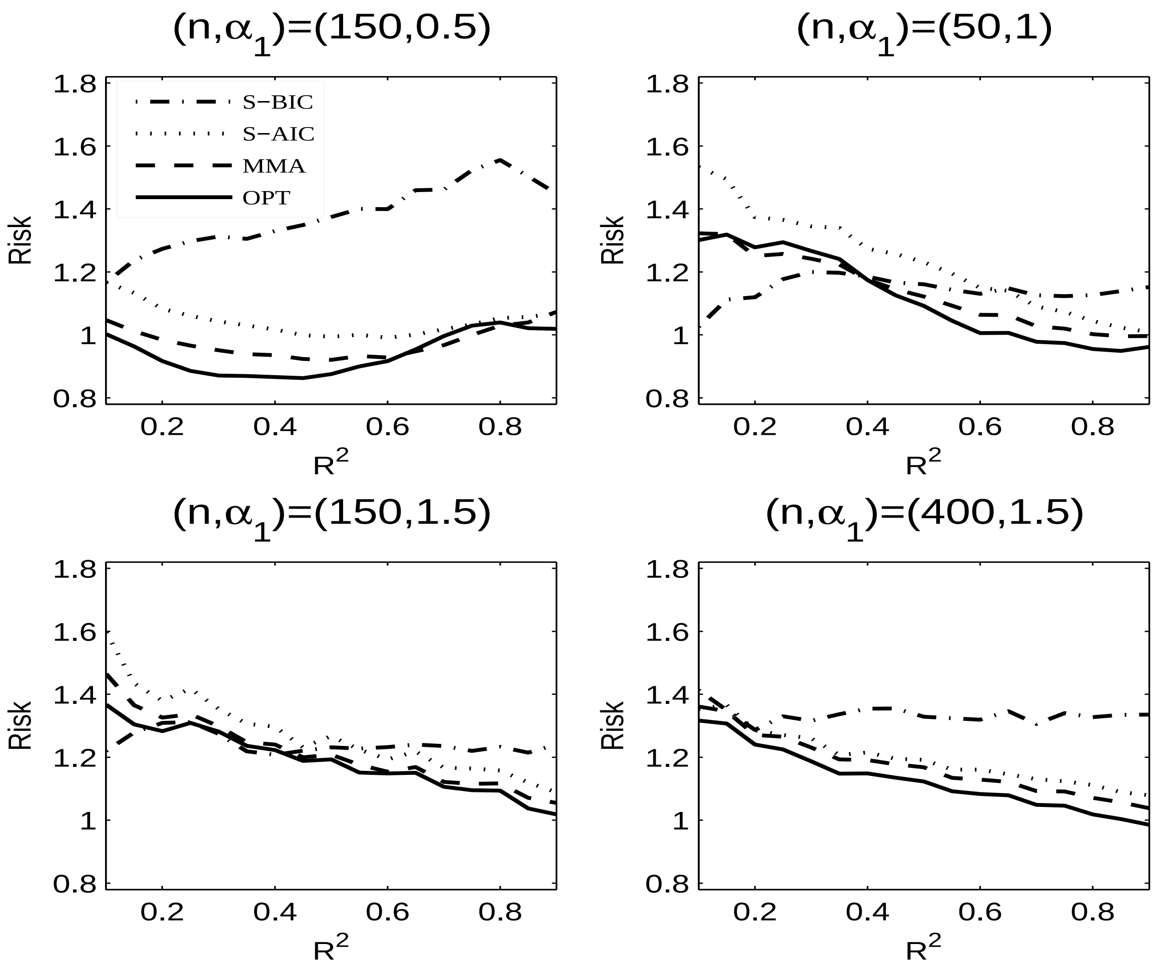

Figure B.3: Results for Example 2: risk comparisons under $L^{(3)}$ loss.

REFERENCES

Carter, R.A.L., Srivastava, M.S., Srivastava, V.K. and Ullah, A. (1990), "Unbiased Estimation of the MSE Matrix of Stein-rule Estimators, Confidence Ellipsoids and Hypothesis Testing," *Econometric Theory*, 6, 63-74.

Giles, D.E.A. and Srivastava, V.K. (1991), "An Unbiased Estimator of the Covariance Matrix of the Mixed Regression Estimator," *Journal of the American Statistical Association*, 86, 441-444.

Hansen, B.E. (2007), "Least Squares Model Averaging," *Econometrica*, 75, 1175-1190.

Whittle, P. (1960), "Bounds for the Moments of Linear and Quadratic Forms in Independent Variables," *Theory of Probability and Its Applications*, 5, 302-305.